

# Supporting Information

## Shedding Light on the Inhibitory Mechanisms of SARS-CoV-1/CoV-2 spike proteins by ACE2-Designed Peptides

Frederico Campos Freitas,<sup>†,¶</sup> Paulo Henrique Borges Ferreira,<sup>†,¶</sup> Denize Cristina Favaro,<sup>‡</sup> and Ronaldo Junio de Oliveira<sup>\*,†</sup>

<sup>†</sup>*Laboratório de Biofísica Teórica, Departamento de Física, Instituto de Ciências Exatas, Naturais e Educação, Universidade Federal do Triângulo Mineiro, Uberaba, MG, Brazil.*

*38064-200*

<sup>‡</sup>*Departamento de Química Orgânica, Instituto de Química, Universidade Estadual de Campinas, São Paulo, SP, Brazil. 13083-970*

<sup>¶</sup>*These authors contributed equally.*

E-mail: ronaldo.oliveira@uftm.edu.br

# List of Figures

S1	Sequence comparison of SARS-CoV-1 and SARS-CoV-2 spike monomers. . .	3
S2	Structure alignment of SARS-CoV-1 and SARS-CoV-2 spike protein RBDs. .	4
S3	Mutational frustration index of the interface contacts between SARS-CoV-1 and 2 spike proteins and the inhibitors <b>1</b> , <b>2</b> , and <b>3</b> . . . . .	6
S4	Configurational frustration index of the interface contacts between SARS- CoV-1 and 2 spike proteins and the inhibitors <b>1</b> , <b>2</b> , and <b>3</b> . . . . .	7
S5	Specific heat as a function of temperature and protein-peptide dimer interface contacts analysis. . . . .	11
S6	Free-energy ( $F$ ) of ACE2 peptides binding to SARS-CoV-1 and SARS-CoV-2 spike-RBD proteins. . . . .	12
S7	Free-energy profiles as a function of the fraction of native contacts ( $F(Q_{peptide})$ )	13
S8	Two-dimensional free-energy ( $F$ ) of SARS-CoV-1 spike RBD bound to ACE2 peptides <b>1</b> , <b>2</b> , and <b>3</b> . . . . .	14
S9	Two-dimensional free-energy ( $F$ ) of SARS-CoV-2 spike RBD bound to ACE2 peptides <b>1</b> , <b>2</b> , and <b>3</b> . . . . .	15
S10	Root mean square fluctuation of SARS-CoV-2 spike RBD bound to ACE2 peptides <b>1</b> , <b>2</b> , and <b>3</b> . . . . .	16
S11	Contact probability as a function of the number of native contacts formed for SARS-CoV-1 spike and three chosen peptides. . . . .	18
S12	Contact probability as a function of the number of native contacts formed for SARS-CoV-2 spike and three chosen peptides. . . . .	19
S13	Hydrophobicity surfaces of the two coronavirus dimers CoV1+ <b>2</b> and CoV2+ <b>3</b> .	20

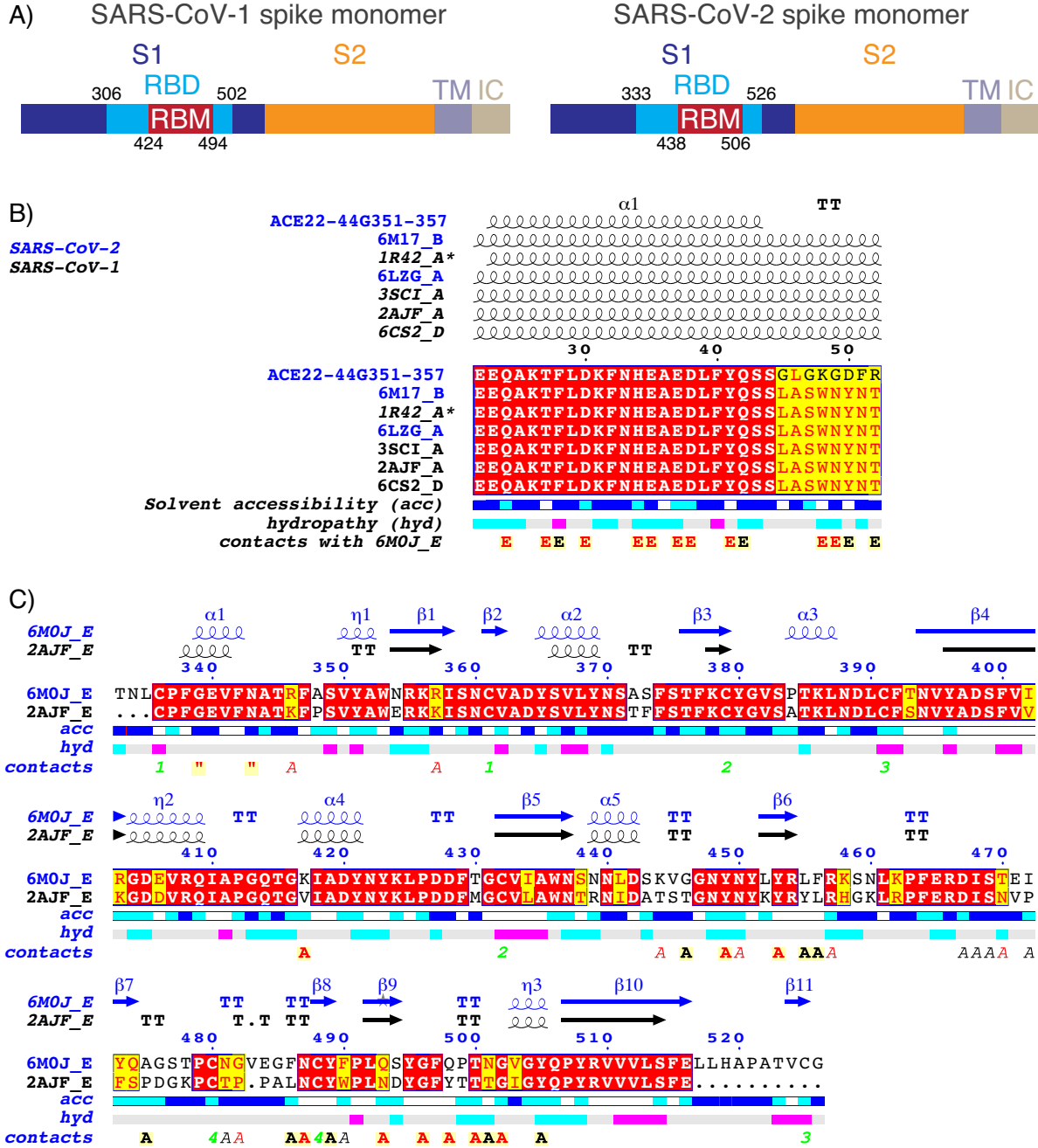


Figure S1: A) Topology of SARS-CoV-1 and SARS-CoV-2 spike protein monomers with domains S1, S2, transmembrane (TM), and intracellular (IC). S1 contains the simulated spike receptor-binding domain (RBD) that is bound to the human ACE2 by the spike receptor-binding motif (RBM). It is also indicated the simulated amino acid residue sequence of both spike RBD. Sequence alignment B) of the designed peptide ACE22-44G351-357 with  $\alpha_1$ -helix of human ACE2 bound with various coronavirus spike protein (except for 1R42 that is absent), and C) of the coronavirus spike protein RBDs. Simulated dimers SARS-CoV-1+ACE22-57 corresponds to PDB code 2AJF and SARS-CoV-2+ACE22-44G351-357 corresponds to 6M0J or 6M17. Alignments were identified by ENDscript 2<sup>1</sup>.

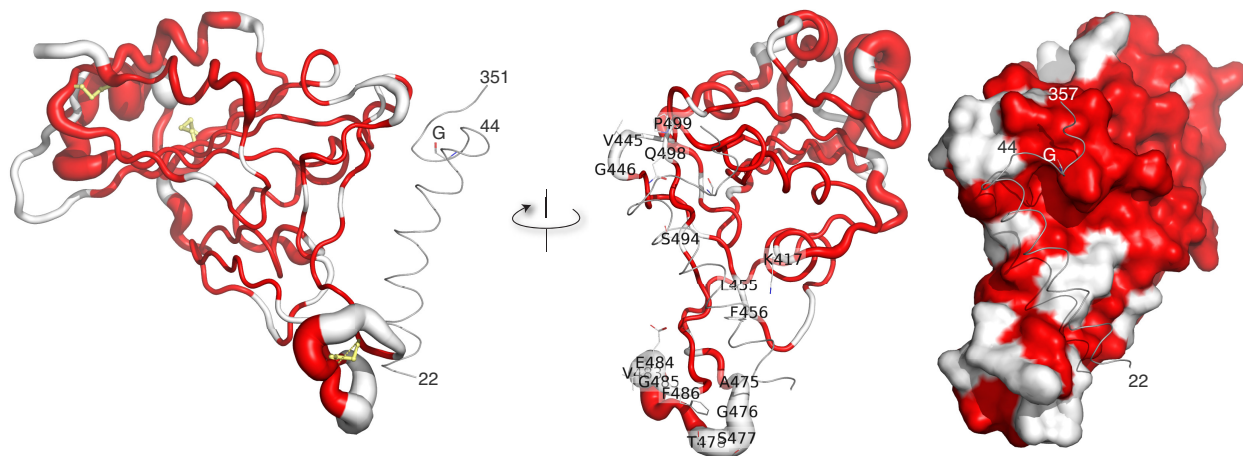


Figure S2: Structure alignment of SARS-CoV-1 and SARS-CoV-2 spike protein RBDs colored according to sequence conservation (red is conserved). The variable tube depiction of the protein main chain is the radius proportional to the differences in  $C_{\alpha}$  between the two homologous protein structures (left and center panels). A surface representation is in the right panel. The thin grey cartoon is the designed peptide ACE22-44G351-357 included in as a reference for visualization of the critical protein contact residues not preserved in the novel coronavirus SARS-CoV-2. Alignment was identified by ENDscript 2 web server<sup>1</sup>.



# Local frustration analysis

Local frustration is evaluated by a Z-score index, which is computed by performing a set of perturbations in a given local of the protein (decoys)<sup>2</sup>. When a pair of residues in contact is perturbed by changing only their identities, we have a mutational frustration index. When the same pair have changed identity and also displaced location, we have a configurational frustration index. The frustration index  $F$  of a contact pair with respect to a set of  $N$  decoys is given by<sup>3</sup>

$$F_{ij} = \frac{E_{ij}^N - \langle E_{i'j'}^U \rangle}{\sqrt{\frac{1}{N} \sum_{k=1}^n (E_{i'j'}^U - \langle E_{i'j'}^U \rangle)^2}} \quad (\text{S1})$$

where  $E^N$  is the native energy and  $\langle E^U \rangle$  is the mean energy of the decoys. A contact with  $F_{ij} < -1$  is recognized as highly frustrated. If  $F_{ij} \geq 0.78$ , the contact is considered minimally frustrated.<sup>2</sup> The contact is neutral when  $F_{ij}$  is between these values. Mutational and configurational frustration indexes for SARS-CoV-1 and 2 spike proteins complexed to the structures of inhibitors **1**, **2** and **3** were calculated with Frustratometer 2 Web server, taking into account the electrostatic interactions in the energy potential (<http://frustratometer.qb.fcen.uba.ar>).<sup>4-7</sup>

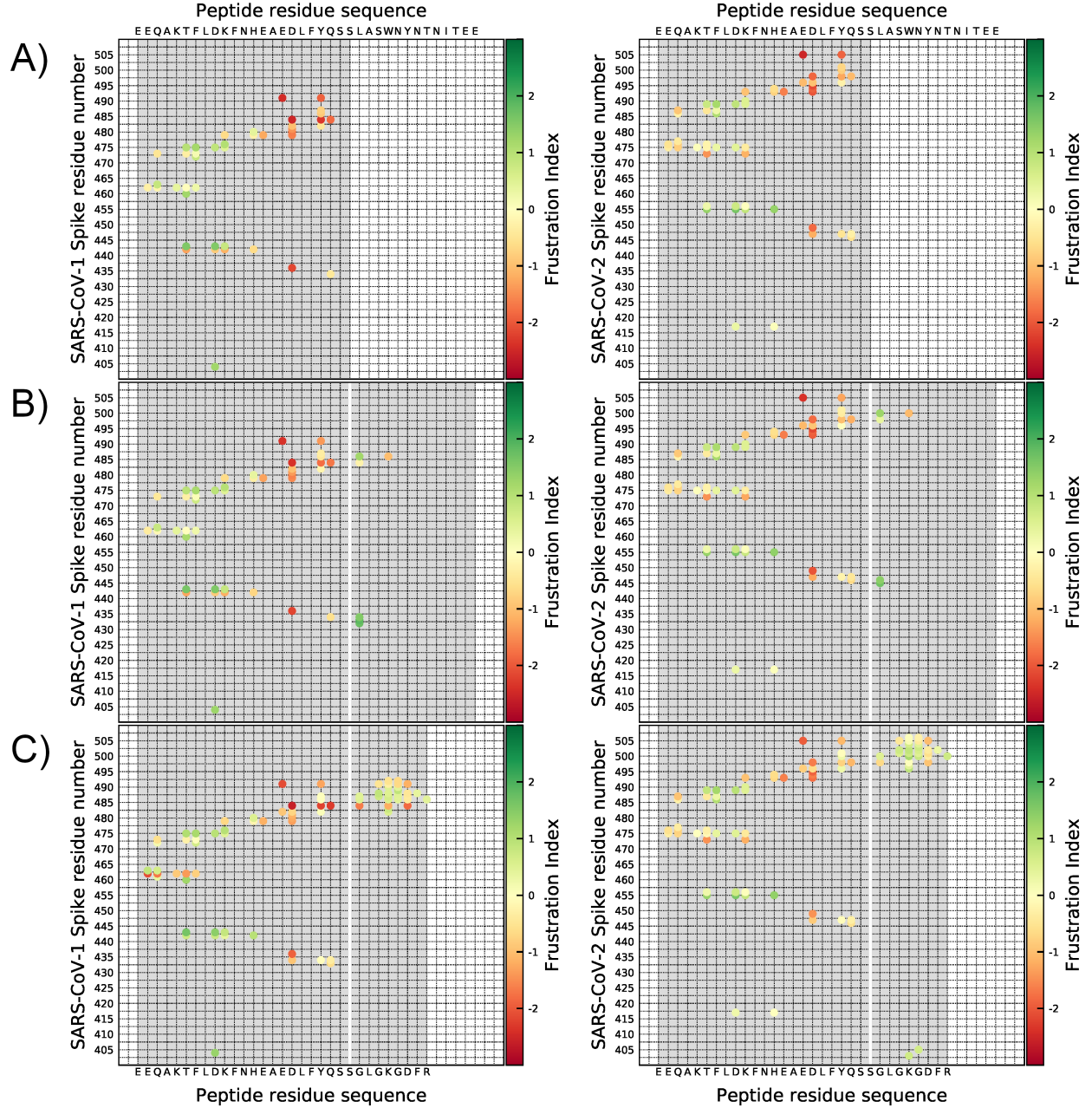


Figure S3: Mutational frustration index of the interface contacts between SARS-CoV-1 and 2 spike proteins and the inhibitors A) **1**, B) **2**, and C) **3**. The peptide residue sequences are highlighted on gray background. The white line in B) and C) divides the original peptide residue sequence of inhibitor **1** from the additional residues in inhibitors **2** and **3**.

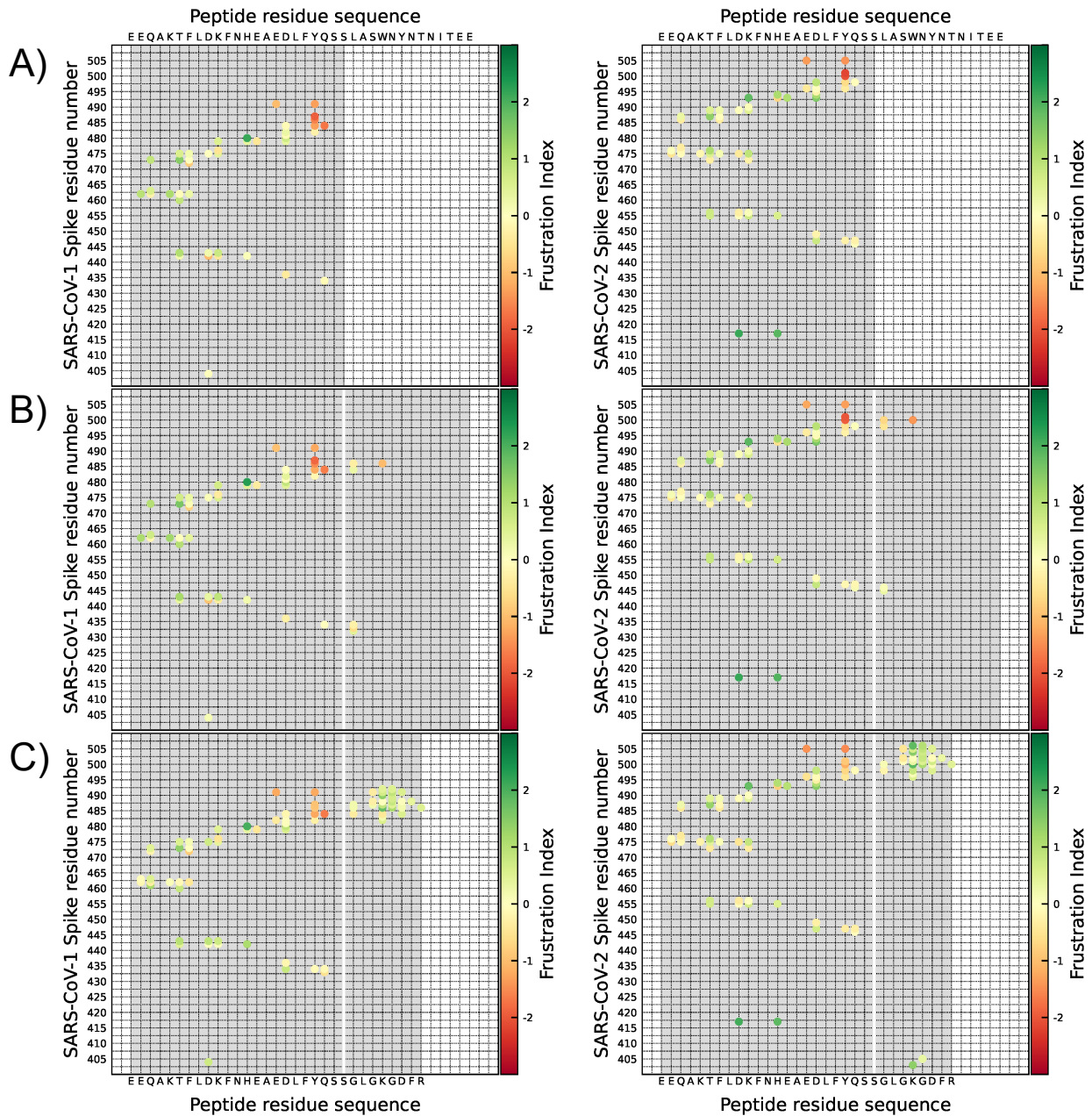


Figure S4: Configurational frustration index of the interface contacts between SARS-CoV-1 and 2 spike proteins and the inhibitors A) 1, B) 2, and C) 3. The peptide residue sequences are highlighted on gray background. The white line in B) and C) divides the original peptide residue sequence of inhibitor 1 from the additional residues in inhibitors 2 and 3.

## Structure-Based Model

A standard all-atom Structure-Based Model was used in this work, in which all non-hydrogen atoms were explicitly represented as beads of unit mass<sup>8</sup>. Bond lengths, bond angles, improper and planar dihedrals were maintained by harmonic potentials. Cosine terms were assigned to flexible dihedral angles. The minima in all these cases were assigned the structure given values. Nonbonded atom pairs in contact were defined between residues  $i$  and  $j$ , where  $i > j + 3$ , using Shadow Contact Map algorithm<sup>9</sup>, with 6Å cut-off distance and 1Å shadowing radius. A Lennard-Jones potential is used to describe these contacts, and all other nonlocal interactions are repulsive. The functional form of the potential is:

$$\begin{aligned}
 V = & \sum_{bonds} \frac{\epsilon_r}{2} (r_i - r_{i,0})^2 + \sum_{angles} \frac{\epsilon_\theta}{2} (\theta_i - \theta_{i,0})^2 + \\
 & + \sum_{impropers} \frac{\epsilon_{\chi_{imp}}}{2} (\chi_i - \chi_{i,0})^2 + \sum_{planar} \frac{\epsilon_{\chi_{planar}}}{2} (\chi_i - \chi_{i,0})^2 + \\
 & + \sum_{backbone} \epsilon_{bb} F_D (\phi_i - \phi_{i,0}) + \sum_{sidechains} \epsilon_{sc} F_D (\phi_i - \phi_{i,0}) + \\
 & + \sum_{contacts} \epsilon_C \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{non-contacts} \epsilon_{nc} \left( \frac{\sigma_{nc}}{r_{ij}} \right)^{12} \quad (S2)
 \end{aligned}$$

where  $F_D (\phi_i - \phi_{i,0}) = [1 - \cos (\phi_i - \phi_{i,0})] + \frac{1}{2}[1 - \cos (3 (\phi_i - \phi_{i,0}))]$  and  $r_{i,0}$ ,  $\theta_{i,0}$ ,  $\chi_{i,0}$  and,  $\phi_{i,0}$  are determined by the given structure. In this model, the parameters used were set to:  $\epsilon_r = 100\epsilon/\text{\AA}^2$ ,  $\epsilon_\theta = 80\epsilon/\text{rad}^2$ ,  $\epsilon_{\chi_{imp}} = 10\epsilon/\text{rad}^2$ ,  $\epsilon_{\chi_{planar}} = 40\epsilon/\text{rad}^2$ ,  $\epsilon_{nc} = 0.1\epsilon$ ,  $\sigma_{nc} = 2.5\text{\AA}$  and,  $\epsilon = 1$ . The constant  $\epsilon$  is in the reduced energy unit and has the value 1 in the topology file, equivalent to  $1\text{kJ/mol}$  if interpreted in Gromacs units. Dihedrals were reweighted accordingly the number of dihedral angles that have a common middle bond ( $N_D$ ), scaling the energy of each dihedral associated with a bond by the factor  $1/N_D$ . The contact and

dihedral energies were divided according to:

$$\frac{\epsilon_{bb}}{\epsilon_{sc}} = 2$$

$$\frac{\sum \epsilon_c}{\sum \epsilon_{bb} + \sum \epsilon_{sc}} = 2 \text{ and}$$

$$\sum \epsilon_c + \sum \epsilon_{bb} + \sum \epsilon_{sc} = N\epsilon$$

with  $N$  being the total number of atoms in the model. As discussed in<sup>10</sup>, these settings have allowed a timestep of 0.002 in reduced units. In this work, it was simulated  $5 \cdot 10^8$  timesteps, writing the coordinates at every 2000 steps.

It was simulated a wide range of temperatures (in reduced units) to obtain the binding temperature. There is not a direct connection between the simulated temperatures and the real-life temperature in which the experiments are performed, since the model is coarse-grained and the target temperature should be around the unit. In the case of the Gromacs package used here, reduced units are around the inverse of the Boltzman constant ( $1/k_B \sim 120$ ) in which temperatures of coarse-grained models should be targeted to 120. Thus, the dimers were extensively simulated from very low temperatures where the system samples only native-bound states to high temperatures where the protein-peptide is always dissociated. We found a curve for  $C_v$  with two peaks for all the simulated dimers. We observed that the first peak corresponds to temperatures below 100 where dissociation of the protein-peptide occurs whereas the spike protein remains relatively folded, and a second peak for temperatures over 120 where the dimer is always unbound and the spike protein completely unfolds. We focus our attention on the dimerization mechanism, since sampling of folding-unfolding of the spike protein was very computationally costly with a time-scale of orders of magnitude larger than the bind-unbound mechanism. So, we defined  $T_{bind}$  as the reference temperature for comparisons of the different simulated dimers by inspecting  $C_v$  that were obtained by differentiating internal energy with respect to the reduced temperature. The

peak of  $C_v$  describes the transition of the system, from low temperatures, which we have a bound state, to high temperatures, which we have an unbound state. We defined  $T_{bind}$  as the transition temperature in a similar manner as the protein folding community defines the transition temperature separating folded and unfolded states,  $T_{fold}$ .

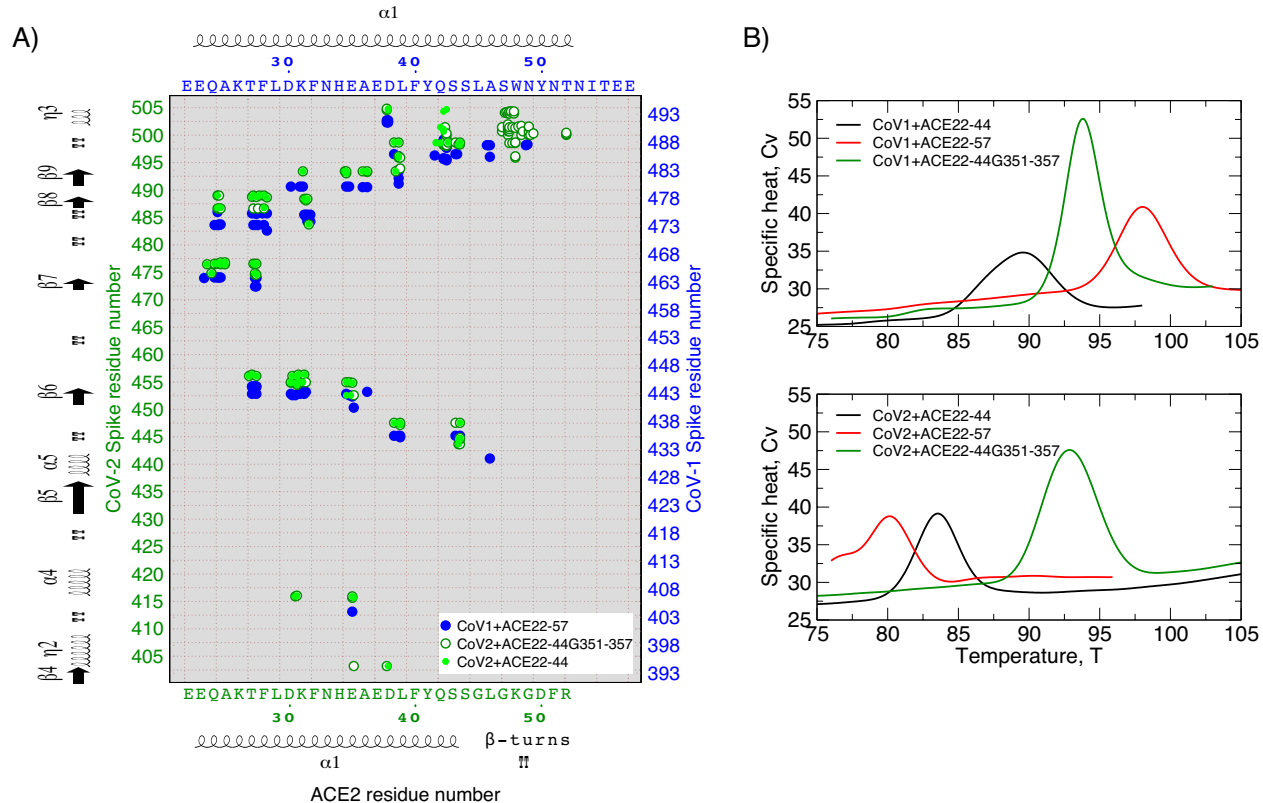


Figure S5: A) Contacts of only the protein-peptide dimer interfaces identified by Shadow Map algorithm<sup>9</sup> and used in the all heavy (non-hydrogen) atoms structure-based model (SBM) simulations. B) Specific heat ( $C_v$ ) as a function of temperature of the dimers simulated with SBM. The peak of each curve corresponds to the critical transition temperature ( $T_{bind}$ ) between bound (low temperatures) and unbound (high temperatures) states of each simulated dimer.

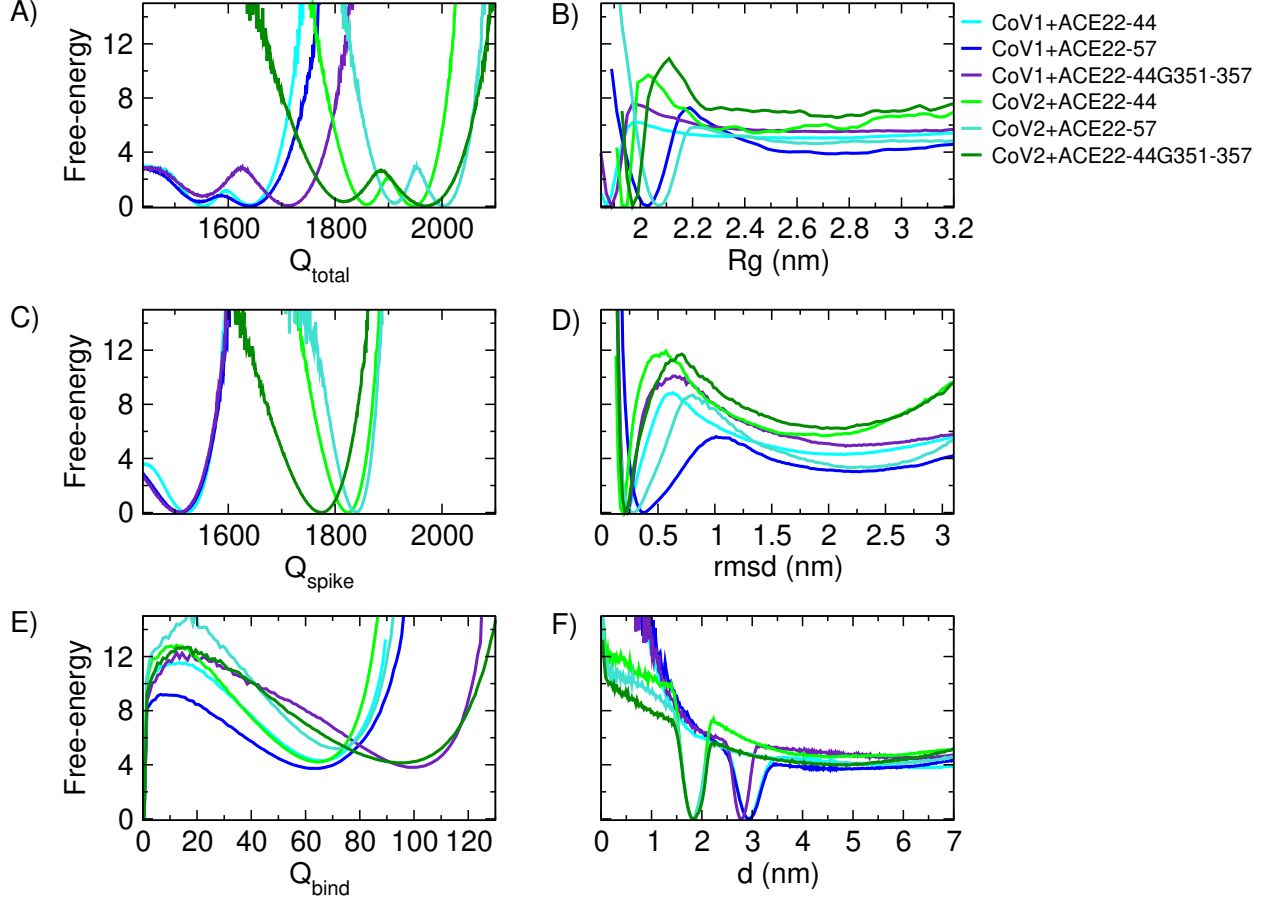


Figure S6: Free-energy ( $F$ ) of ACE2 peptides binding to SARS-CoV-1 and SARS-CoV-2 spike-RBD proteins.  $F$  profiles are shown as a function of reaction coordinates A) inter plus intrachain contacts ( $Q_{total}$ ), B) radius of gyration ( $Rg$ ), C) intrachain protein contacts ( $Q_{spike}$ ), D) root means square deviation (rmsd), E) native peptide-protein chain interface contacts ( $Q_{bind}$ ) and F) distance between center of mass of the protein-peptide dimer ( $d$ ). The designed peptide ACE22-44G351-357 and the sort  $\alpha_1$  peptide (ACE22-44) binds to SARS-CoV-2 protein with higher  $F$  barrier when comparing with the  $\alpha_1$  peptide (ACE22-57) binding to SARS-CoV-1 spike domain. Also, for SARS-CoV-2, the designed ACE22-44G351-357 has higher binding  $F$  barrier than the sort ACE22-44 in some reaction coordinates.  $F$  profiles are in units of  $k_B T_{bind}$  with  $T_{bind}$  been the respective dimer binding temperature.



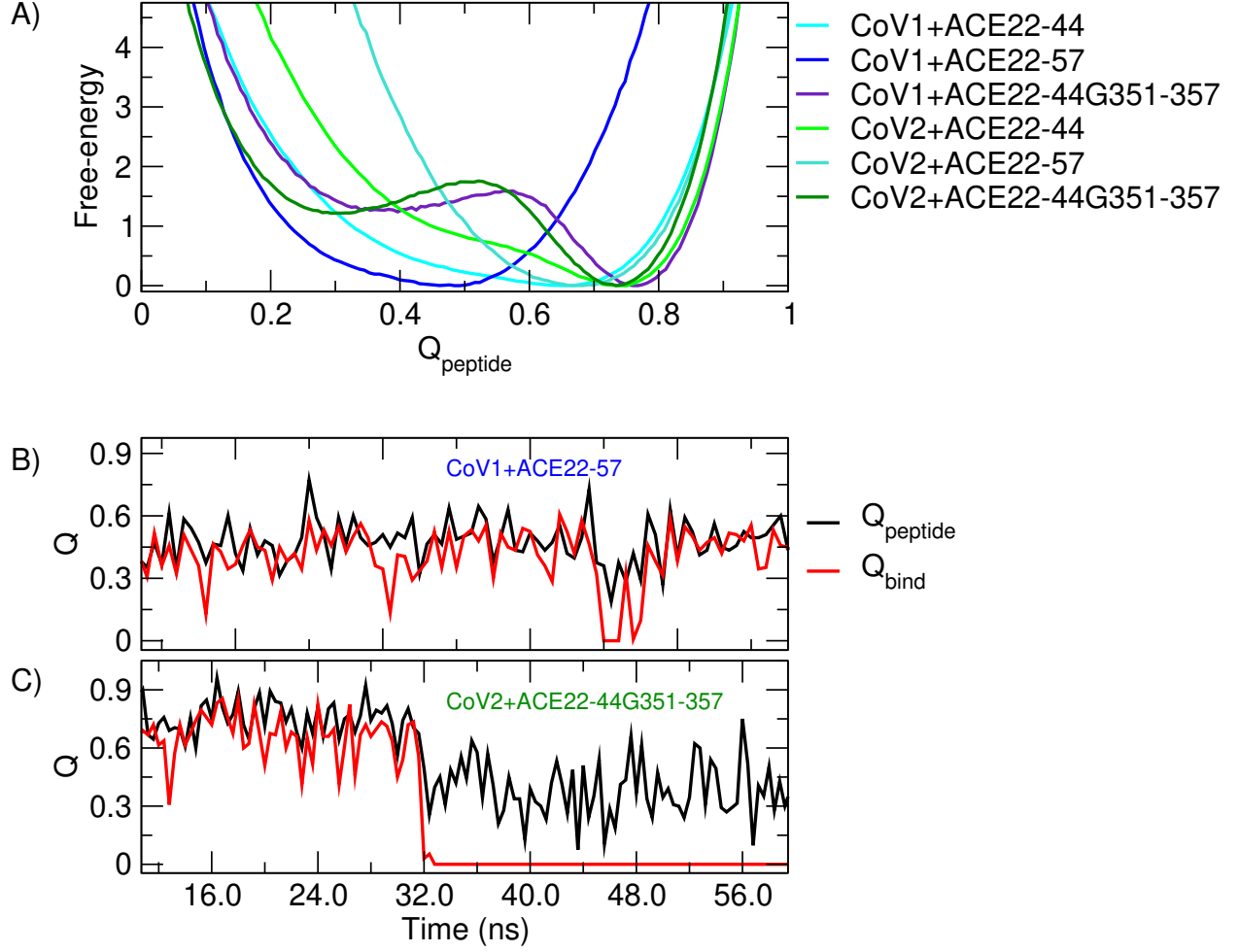


Figure S7: A) Free-energy profiles as a function of the fraction of native contacts ( $F(Q_{\text{peptide}})$ ) of ACE2-based peptides in the binding-folding simulations at  $T_{\text{bind}}$ . Time trajectories of the reaction coordinates fraction of native peptide contacts ( $Q_{\text{peptide}}$  in black) and protein-peptide binding contacts ( $Q_{\text{bind}}$  in red) for dimers B) CoV1+**2** and C) CoV2+**3** at  $T_{\text{bind}}$ .  $F$  profiles are in units of  $k_B T_{\text{bind}}$  with  $T_{\text{bind}}$  been the respective system binding temperature.

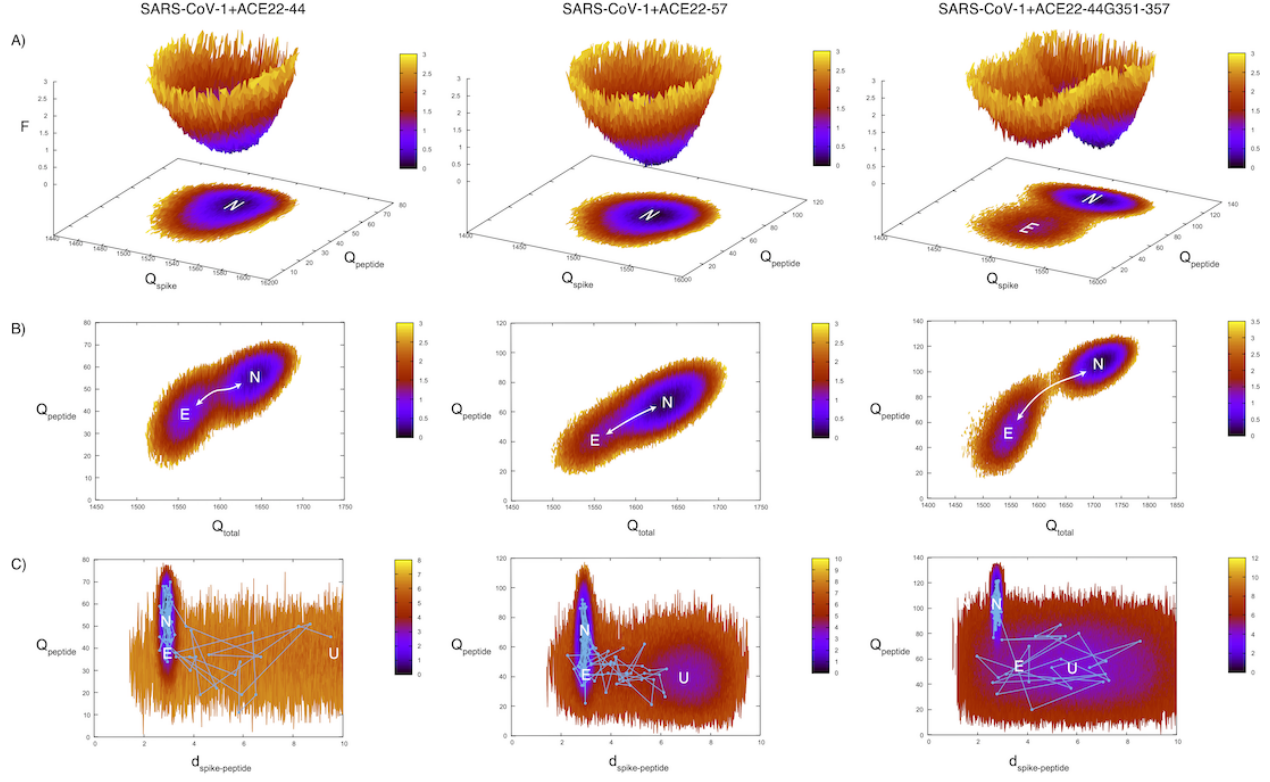
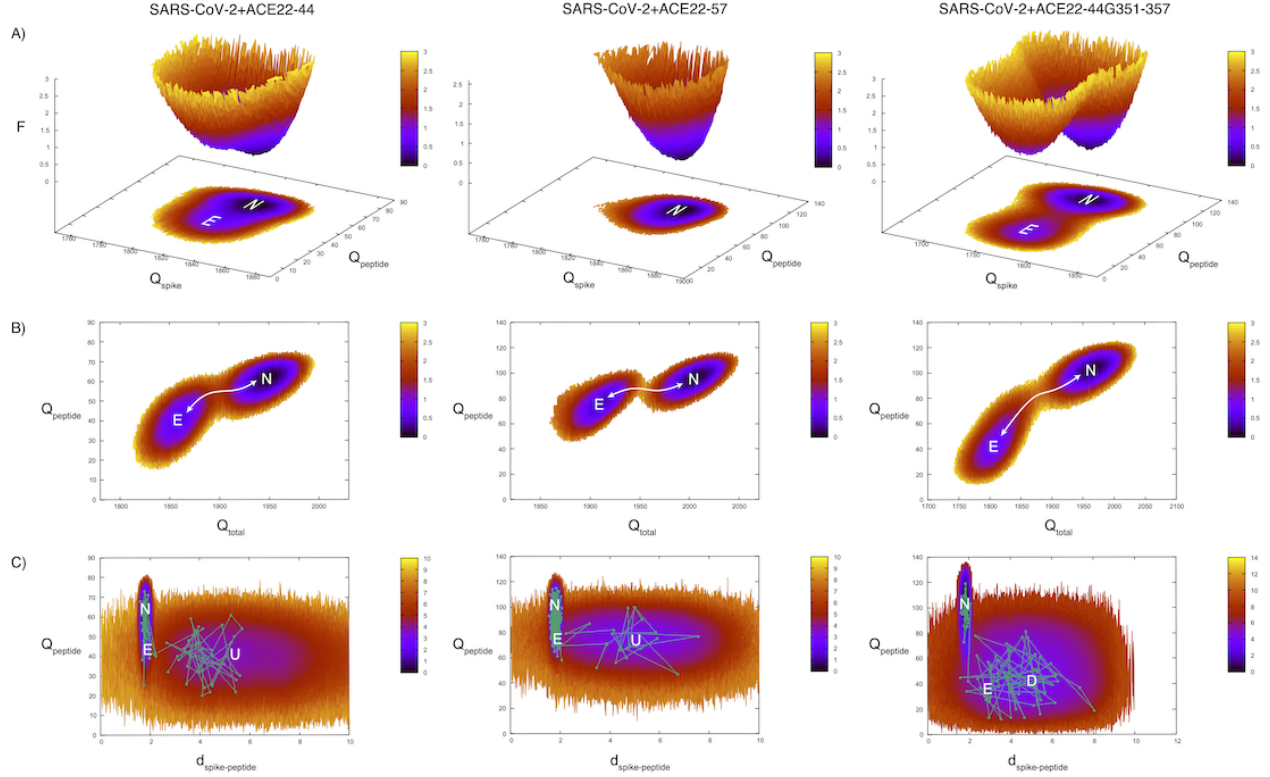


Figure S8: Two-dimensional free-energy ( $F$ ) of SARS-CoV-1 spike RBD bound to ACE2 peptides **1** (left panels), **2** (center panels) and **3** (right panels). N is the native state and E is the encounter complex state.  $F$  profiles are shown as a function of reaction coordinates of native A) peptide ( $Q_{peptide}$ ) and protein ( $Q_{spike}$ ) contacts and B) inter plus intrachain contacts ( $Q_{total}$ ) and  $Q_{peptide}$ . C)  $F$  profiles as a function of the reaction coordinate distance between center of mass of the protein-peptide dimer ( $d_{spike-peptide}$ ). The white arrows in B) and C) represent the average folding/binding mechanism between the two N and E states. The connected points in C) are illustrative unfolding/unbinding trajectories from N to D well.  $F$  profiles are in units of  $k_B T_{bind}$  with  $T_{bind}$  been the respective system binding temperature.



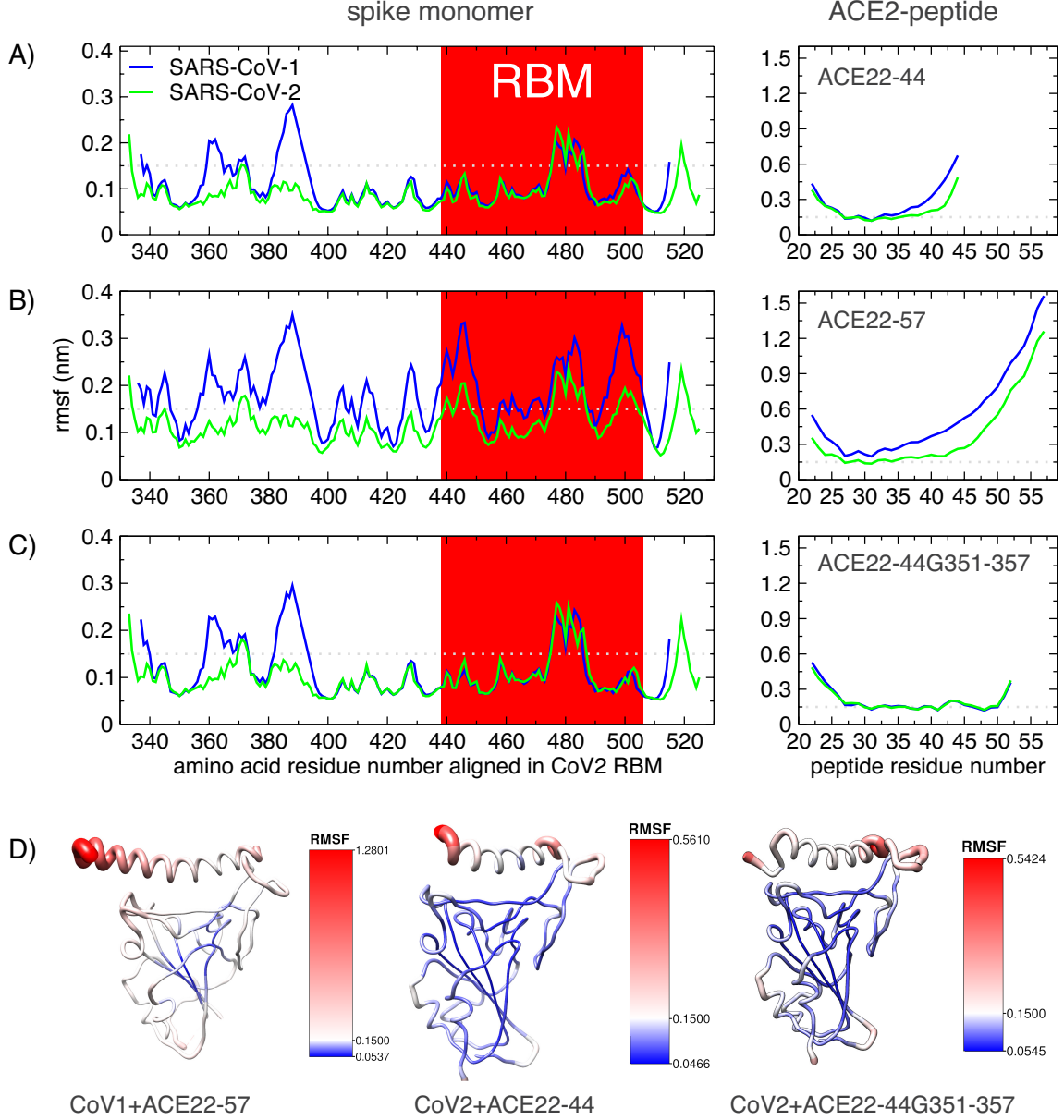


Figure S10: Root mean square fluctuation (rmsf) of SARS-CoV-2 spike RBD (left panels in A-C) bound to ACE2 peptides A) **1**, B) **2** and C) **3**, respectively, in the right panels. The grey dashed line is a guide to the eye at  $\text{rmsf} = 0.15 \text{ \AA}$ . D) It is show selected protein-peptide rmsfs embedded into their native structures with the tube radius and color been the respective rmsfs in A-C. Both, protein and peptide, has the highest residue position fluctuations in SARS-CoV-1 on average. For SARS-CoV-2, the random coil region around the RBM residue 480 (kind of a “claw” domain in the structure), also has high fluctuations (over the rmsf guide line) comparing with other parts of the structure that are mostly bellow the guide line. The temperature selected to calculate each rmsf was that before the binding transition begins in  $C_v(T)$  curve of each protein-peptide dimer, which was around  $0.95T_b$  of each simulation.

# Contact Probability along reaction coordinate

To evaluate how stable are the contacts made by each atom or residue during the conformational changes undergone by the simulated system, one can calculate the contact formation probability for each atom given the total number of contacts made. This idea can be summarized by:

$$p(i, Q) = \sum_j \frac{1}{N_j} p_{ij}(Q) \quad (\text{S3})$$

where  $p(i, Q)$  is the probability of a native contact being formed by the atom  $i$  given the total number of formed contacts  $Q$ . A native contact was counted when the distance of a given pair of atoms was equal or smaller than 1.5 times the distance found in the structure. The pairs in contact were determined using the Shadow Contact Map algorithm<sup>9</sup> with standard settings, as aforementioned. For each simulated system, the set of contacts was divided into three subsets: the contacts between atoms of the same structure (either peptide or spike) and the contacts between spike and peptide atoms, called bind contacts. In figure S11 is shown the probability of a contact being formed for SARS-CoV-1 case as a function of total number of native contacts formed for each atom. The same is shown in figure S12 for SARS-CoV-2. To a better visualization, the atom number was replaced by its correspondent residue number. In S11A and in S12A the contact probability is compared among the bind contacts ( $Q_{bind}$ ) between the spike protein and the three chosen peptides. The comparison between the internal peptide and spike contacts can be seen in figures S11B and S11C for CoV-1 and, S12B and S12C for CoV-2, respectively. All contact probability maps were generated using a python script freely available at <https://github.com/ronaldolab/pyContactProbability>.

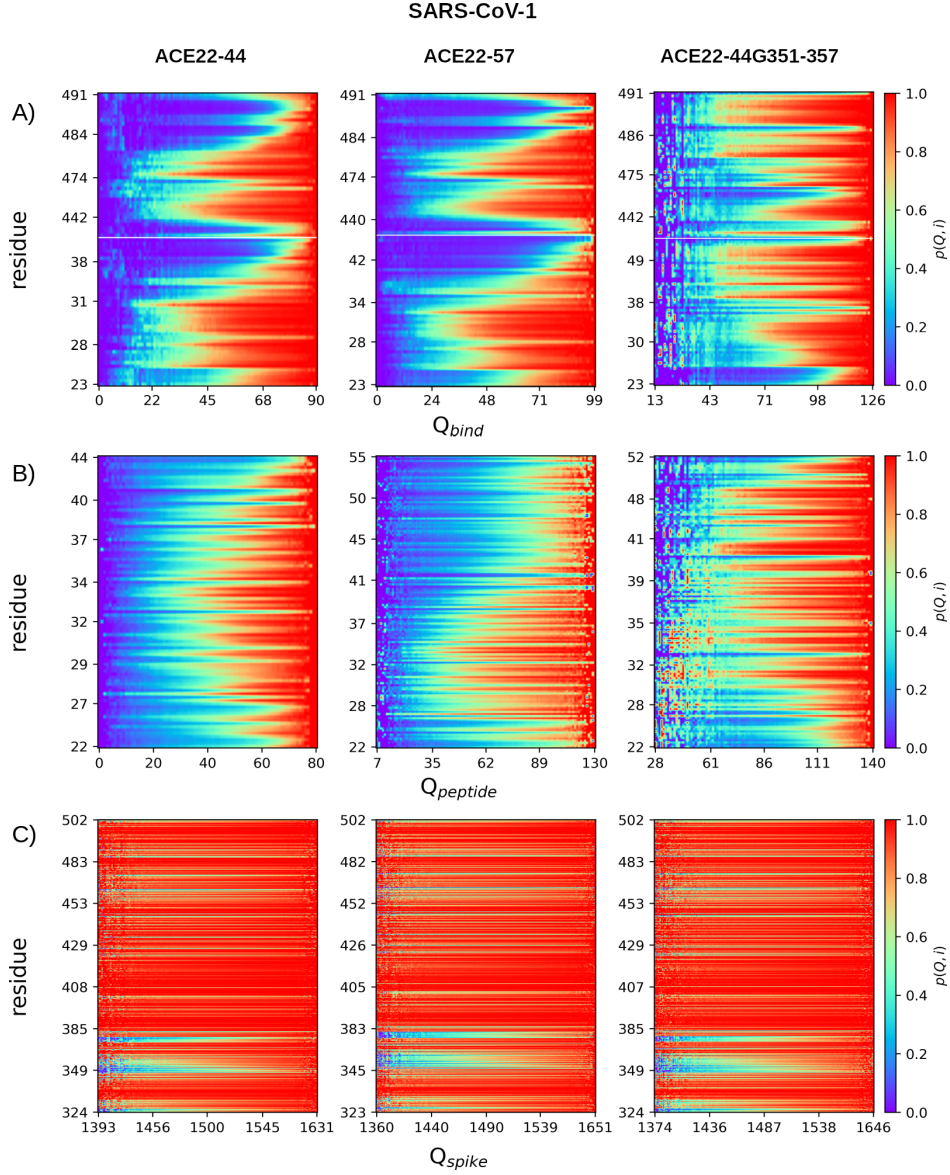


Figure S11: Contact probability as a function of the number of native contacts formed for the SARS-CoV-1 spike and three chosen peptides. On the left column, the plots are related to ACE22-57 (inhibitor **1**), in the middle column they are associated to ACE22-44 (inhibitor **2**) and, on the right, the plots are associated to ACE22-44G351-357 (inhibitor **3**). A) Comparison between the probability of a native contact being formed as a function of the number of binding contacts between the spike and the peptide ( $Q_{bind}$ ). There is a small gap between the peptide and the spike residues. In B) the contact probability as a function of the internal peptide native contacts ( $Q_{peptide}$ ) are compared for the three chosen peptides. C) The contact probability as a function of the internal native spike contacts are shown for the three chosen cases. The probability is high for all three analyzed cases, for all residues, indicating the spike stability during the binding events. All the plots were generated at  $T_{bind}$  for each case.



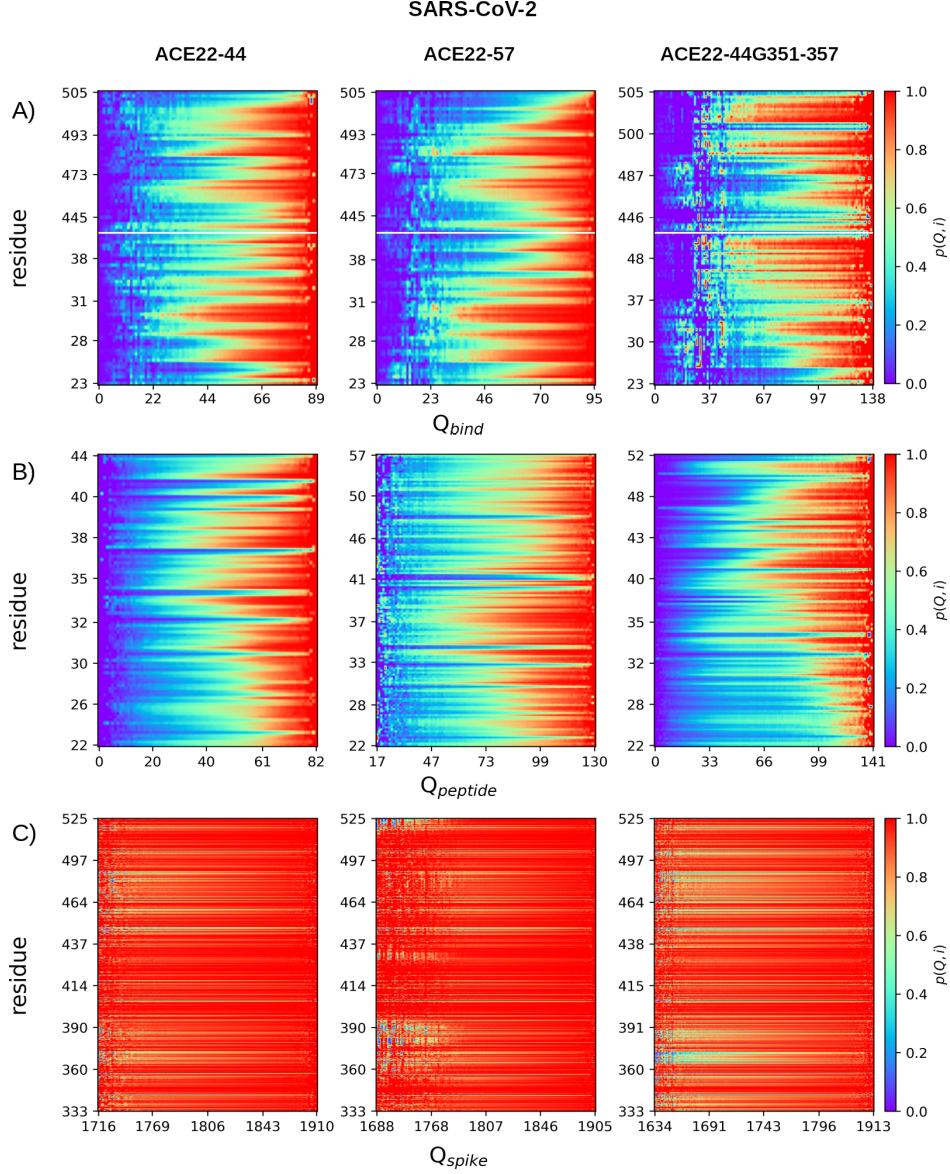


Figure S12: Contact probability as a function of the number of native contacts formed for the SARS-CoV-2 spike and three chosen peptides. On the left column, the plots are related to ACE22-57 (inhibitor **1**), in the middle column they are associated to ACE22-44 (inhibitor **2**) and, on the right, the plots are associated to ACE22-44G351-357 (inhibitor **3**). A) Comparison between the probability of a native contact being formed as a function of the number of binding contacts between the spike and the peptide ( $Q_{bind}$ ). There is a small gap between the peptide and the spike residues. In B) the contact probability as a function of the internal peptide native contacts ( $Q_{peptide}$ ) are compared for the three chosen peptides. C) The contact probability as a function of the internal native spike contacts are shown for the three chosen cases. The probability is high for all three analyzed cases, for all residues, indicating the spike stability during the binding events. All the plots were generated at  $T_{bind}$  for each case.

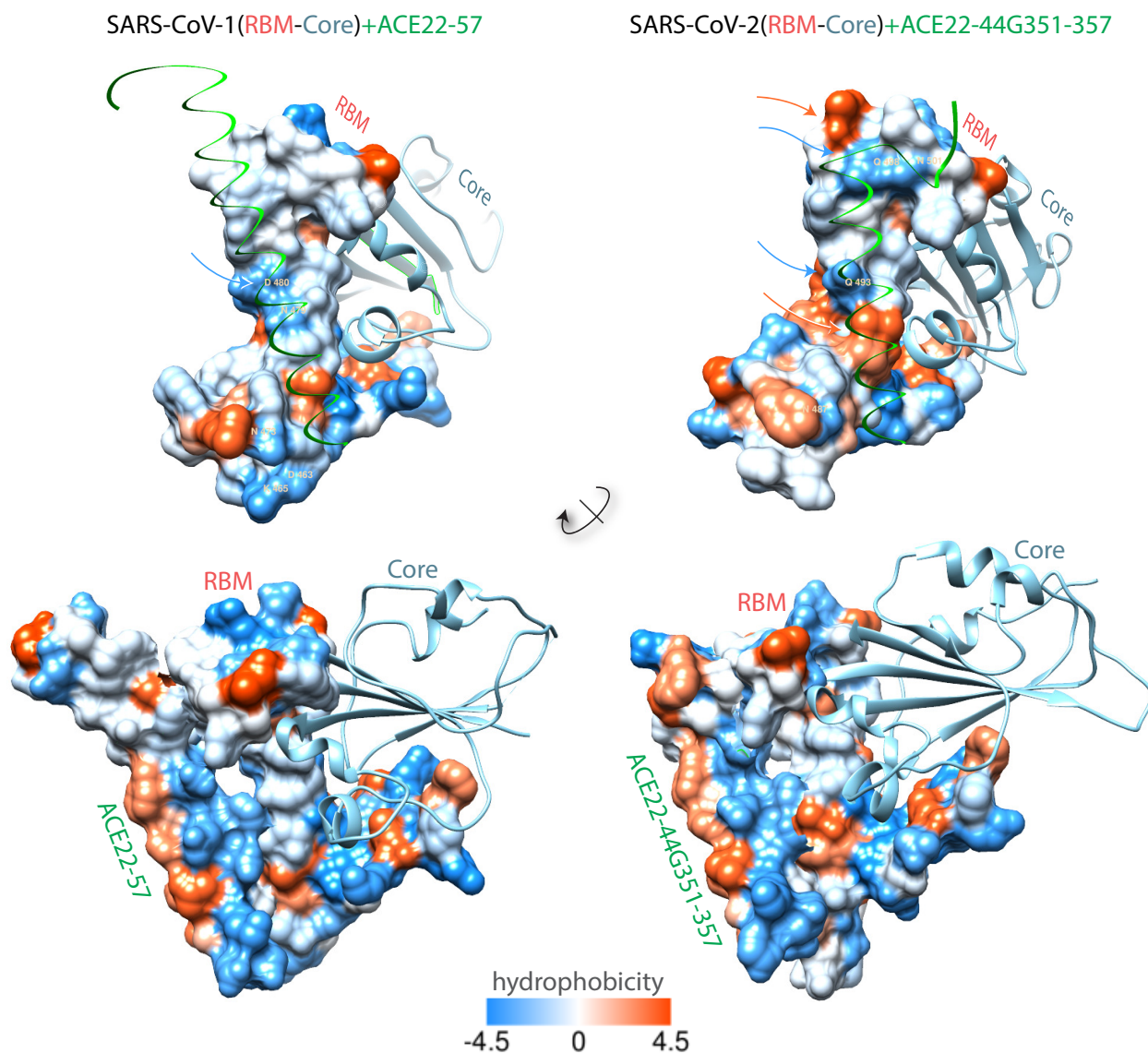


Figure S13: Hydrophobicity surfaces of the two coronavirus dimers CoV1+**2** (left panels) and CoV2+**3** (right panels). The hydrophobicity surfaces are colored from blue for the most hydrophilic, to white, to orange-red for the most hydrophobic. Top panels present the hydrophobicity surfaces only for the spike protein RBMs and lower panels also show the surface for the inhibitor peptides. Spike protein cores are shown as ribbons. CoV2 RBM has two binding regions more hydrophobic (orange arrows) and one more hydrophilic (blue arrow around residues Q498 and N501) than CoV1.



## References

- (1) Robert, X.; Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014**, *42*, W320–W324.
- (2) Ferreiro, D.; Hegler, J.; Komives, E.; Wolynes, P. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19819–19824.
- (3) Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G. Frustration in biomolecules. *Q. Rev. Biophys.* **2014**, *47*, 285–363.
- (4) Jenik, M.; Parra, R. G.; Radusky, L. G.; Turjanski, A.; Wolynes, P. G.; Ferreiro, D. U. Protein frustratometer: a tool to localize energetic frustration in protein molecules. *Nucleic Acids Res.* **2012**, *40*, W348–W351.
- (5) Parra, R. G.; Schafer, N. P.; Radusky, L. G.; Tsai, M.-Y.; Guzovsky, A. B.; Wolynes, P. G.; Ferreiro, D. U. Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res.* **2016**, *44*, W356–W360.
- (6) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J. Phys. Chem. B* **2012**, *116*, 8494–8503.
- (7) Tsai, M.-Y.; Zheng, W.; Balamurugan, D.; Schafer, N. P.; Kim, B. L.; Cheung, M. S.; Wolynes, P. G. Electrostatics, Structure Prediction, and the Energy Landscapes for Protein Folding and Binding. *Prot. Sci.* **2016**, *25*, 255–269.
- (8) Whitford, P.; Noel, J.; Gosavi, S.; Schug, A.; Sanbonmatsu, K.; Onuchic, J. An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins: Struct. Func. Bioinf.* **2009**, *75*, 430–441.

- (9) Noel, J. K.; Whitford, P. C.; Onuchic, J. N. The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function. *J. Phys. Chem. B* **2012**, *116*, 8692–8702.
- (10) Noel, J. K.; Levi, M.; Raghunathan, M.; Lammert, H.; Hayes, R. L.; Onuchic, J. N.; Whitford, P. C. SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLOS Comput Biol* **2016**, *12*, e1004794.